Kybernetika

Tomáš Hobza; Leandro Pardo; Igor Vajda

Robust median estimator for generalized linear models with binary responses

Kybernetika, Vol. 48 (2012), No. 4, 768--794

Persistent URL: http://dml.cz/dmlcz/143059

KYBERNETIKA — VOLUME 48 (2012), NUMBER 4, PAGES 768–794

ROBUST MEDIAN ESTIMATOR FOR GENERALIZED LINEAR MODELS WITH BINARY RESPONSES

Tomáš Hobza, Leandro Pardo and Igor Vajda

The paper investigates generalized linear models (GLM's) with binary responses such as the logistic, probit, log-log, complementary log-log, scobit and power logit models. It introduces a median estimator of the underlying structural parameters of these models based on statistically smoothed binary responses. Consistency and asymptotic normality of this estimator are proved. Examples of derivation of the asymptotic covariance matrix under the above mentioned models are presented. Finally some comments concerning a method called enhancement and robustness of median estimator are given and results of simulation experiment comparing behavior of median estimator with other robust estimators for GLM's known from the literature are reported.

Keywords: generalized linear models, binary responses, statistical smoothing, statistical enhancing, maximum likelihood estimator, median estimator, consistency, asymptotic normality, efficiency, robustness

Classification: 62F10, 62F12, 62F35

1. INTRODUCTION AND BASIC CONCEPTS

Let $Y_1, \ldots, Y_n$ be independent Bernoulli random variables with parameters $\pi_1, \ldots, \pi_n$,

$$Y_1 \sim Be(\pi_1), \ldots \, , Y_n \sim Be(\pi_n).$$

We assume that the Bernoulli parameters $\pi_i = \Pr(Y_i = 1) = EY_i \in (0,1)$ are predictable by means of unknown structural parameters $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0d})^T$ and explanatory variables $\boldsymbol{x}_i^T = (x_{i1}, \ldots, x_{id})$, $1 \leq i \leq n$, through the linear predictor

$$g(\pi_i) = \sum_{j=1}^{d} x_{ij}\beta_{0j}, \quad 1 \leq i \leq n \tag{1}$$

for a given monotone function $g(\pi) : (0,1) \longmapsto \mathbb{R}$. So we consider *generalized linear models* (GLM's) with binary responses and *link functions* $g(\pi)$. In the sequel we restrict ourselves to the strictly monotone and infinitely differentiable link functions and we use the *inverse link functions* $\pi(t) = g^{-1}(t) : \mathbb{R} \longmapsto (0,1)$ which are strictly monotone and infinitely differentiable too and satisfy the relation

$$\pi_i = \pi(\boldsymbol{x}_i^T \boldsymbol{\beta}_0) \stackrel{\Delta}{=} \pi_i(\boldsymbol{\beta}_0), \quad 1 \leq i \leq n. \tag{2}$$

A wide choice of GLM's is available. Some of the most interesting are the following:

a) *Logistic model.* In this model $\pi(t)$ is the standard logistic distribution function,

$$\pi(t) = \frac{e^t}{1+e^t} \quad \text{and} \quad g(\pi) = \ln(\pi/(1-\pi)). \tag{3}$$

b) *Probit model.* Here $\pi(t)$ is the standard normal distribution function,

$$\pi(t) = \Phi(t) \quad \text{and} \quad g(\pi) = \Phi^{-1}(\pi). \tag{4}$$

c) *Log-log model.* In the log-log model $\pi(t)$ is the reflected standard Gumbel distribution function $G(-t) = 1 - G(t)$, i. e.,

$$\pi(t) = \exp(-\exp t) \quad \text{and} \quad g(\pi) = -\ln(-\ln \pi). \tag{5}$$

d) *Complementary Log-log model.* In this model $\pi(t)$ is the standard Gumbel distribution function,

$$\pi(t) = 1 - \exp(-\exp t) \quad \text{and} \quad g(\pi) = \ln(-\ln(1-\pi)). \tag{6}$$

e) *Scobit model.* In this model

$$\pi(t) = 1 - (1+e^t)^{-\lambda} \quad \text{and} \quad g(\pi) = \ln((1-\pi)^{-1/\lambda} - 1), \quad \lambda > 0. \tag{7}$$

f) *Power logit model.* In this model

$$\pi(t) = (1+e^{-t})^{-\lambda} \quad \text{and} \quad g(\pi) = \ln\left(\frac{1}{\pi^{-1/\lambda} - 1}\right), \quad \lambda > 0. \tag{8}$$

For more details about these GLM's see p. 108 in McCullagh and Nelder [15], Menéndez et al. [16] and references therein. For Scobit and Power logit models see Nagler [19] and Prentice [21], respectively.

The main purpose of this paper is the estimation of the parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ in the GLM considered in (1). The maximum likelihood estimator(MLE), $\boldsymbol{\beta}_n = \boldsymbol{\beta}_n(Y_1, \ldots, Y_n)$ is obtained by

$$\boldsymbol{\beta}_n = \arg\min \sum_{i=1}^{n} d_i(\boldsymbol{\beta}) \tag{9}$$

being

$$d_i(\boldsymbol{\beta}) = -Y_i \ln \pi_i(\boldsymbol{\beta}) - (1-Y_i)\ln(1-\pi_i(\boldsymbol{\beta})) \,. \tag{10}$$

In Hobza et al. [9] was established the undesired influence of contaminations in the MLE in Logistic Regression Models. It is easy to observe that the arguments given in Hobza et al. [9] for the logistic regression model are valid to observe the undesired influence of contaminations in the MLE in GLM. In order to overcome the problem many different estimators, for the logistic regression model, have been introduced and studied.

See, Pregibon [20], Morgenthaler [18], Bianco and Yohai [2], Croux and Haesbroeck [3], Kordzakhia et al. [11], Adimari and Ventura [1], Rousseeuw and Christmann [22], Gervini [5], Hobza et al. [9]. The problem in GLM's has not been studied in the same way.

In this paper we propose a new robust M-estimator of the parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ in GLM's with binary responses obtained by application of the classical robust L_1-method (cf. Hampel et al. [6], Jurečková and Sen [10] or Zwanzig [23]) to the continuous responses

$$Z_i = Y_i + U_i, \qquad 1 \le i \le n \tag{11}$$

obtained by adding mutually and on Y_i independent $U(0,1)$-distributed (i. e. uniformly on $(0,1)$ distributed) random variables U_i to the original above introduced binary responses

$$Y_i \sim Be\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right). \tag{12}$$

In other words, we define the estimator

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \left| Z_i - m\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)\right| \tag{13}$$

for Z_i given by (11), (12) and for the median function

$$m(p) = F_p^{-1}(1/2) = \inf\{z \in \mathbb{R} : F_p(z) \ge 1/2\}$$

corresponding to the class of distribution functions F_p of the random variables

$$Z = Be(p) + U(0,1)$$

when the parameter p varies in the interval $(0,1)$. Obviously, for each $p \in (0,1)$ and $z \in \mathbb{R}$

$$F_p(z) = (1-p)\, z\boldsymbol{I}(0 < z \le 1) + (1 - 2p + pz))\,\boldsymbol{I}(1 < z \le 2) + \boldsymbol{I}(z > 2) \tag{14}$$

and the median function has the explicit form

$$m(p) = 1 + \frac{p - 1/2}{p \vee (1-p)}, \quad 0 < p < 1. \tag{15}$$

Here and in the rest of the paper we use the notation

$$a \vee b = \max\{a, b\} \quad \text{and} \quad a \wedge b = \min\{a, b\}.$$

In the following we shall call the estimator defined in (13) by *Med-estimator*.

Median function $m(p)$ defined by the continuous random variables $Z = Be(p) + U(0,1)$ is strictly increasing in $p \in [0,1]$. Since the function $\pi(t)$ is assumed to be strictly monotone, the argument $m\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)$ in (13) detects every change of the product $\boldsymbol{x}^T\boldsymbol{\beta}$. Contrary to this, the median function $\widetilde{m}(p)$ defined in a similar manner by the discrete random variables $Y = Be(p)$ themselves is piecewise constant in $p \in (0,1)$ so that

$\widetilde{m}\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)$ is insensitive to small variations of the product $\boldsymbol{x}^T\boldsymbol{\beta}$. Therefore the robust L_1-estimation cannot be applied directly to the GLM responses Y_i, i.e., the estimator

$$\widetilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left|Y_i - \widetilde{m}\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)\right|$$

is not of a too much practical interest.

In the next section we shall prove the consistency as well as the asymptotic normality of the *Med-estimator*.

2. ASYMPTOTIC THEORY

In this section we study the asymptotics of the *Med-estimator* $\widehat{\boldsymbol{\beta}}_n$ from (13) which estimates the true parameters $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ of the GLM using the statistically smoothed responses

$$Z_i = Y_i + U_i \sim F_{\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)}(z) \qquad \text{(cf. (11))}$$

to the regressors $\boldsymbol{x}_i$ where $\pi : \mathbb{R} \longmapsto (0,1)$ is strictly monotone and infinitely differentiable, and $F_p(z)$ is given by (14).

Our results are based on what Liese and Vajda [12, 13, 14] proved concerning the general median estimators

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left|Z_i - m\left(u\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)\right| \tag{16}$$

for a given function $m : \Theta \mapsto \mathbb{R}$ of parameters $\boldsymbol{\beta}_0$ in the general statistical model

$$Z_i \sim F_{u\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)}(z), \qquad 1 \le i \le n, \tag{17}$$

being $u : \mathbb{R} \to \Theta$ a smooth mapping and $F_\theta, \theta \in \Theta \subset \mathbb{R}$ is a family of distribution functions on $\mathbb{R}$.

We shall study and adapt to the present estimators (13) the following conditions (c1) – (c8) for consistency and asymptotic normality established by these authors.

(c1) The fixed regressors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ are from a compact set $\mathcal{X} \subset \mathbb{R}^d$ and the probability measures

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\boldsymbol{x}_i} \tag{18}$$

tend weakly for $n \to \infty$ to a probability measure Q on Borel subsets of $\mathcal{X}$.

We can observe that if the regressors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are independently generated by a probability measure Q on the Borel subsets of a compact set $\mathcal{X} \subset \mathbb{R}^d$ then $(c1)$ holds almost surely for these $\mathcal{X}$ and Q. For example, if the dimension $d = 1$ then, by the Glivenko theorem, the empirical probability measure (18) tends almost surely to Q in the Kolmogorov distance. But the convergence in this distance implies the weak convergence required by (c1).

(c2) The smallest eigenvalue of the matrix

$$\boldsymbol{\Sigma} = \int_{\mathcal{X}} \boldsymbol{x}\boldsymbol{x}^T \, \mathrm{d}Q(\boldsymbol{x}) \tag{19}$$

is positive. Hence for every $\boldsymbol{\beta} \in \mathbb{R}^d$ different from $\boldsymbol{\beta}_0$

$$Q\left(\boldsymbol{x} \in \mathcal{X} : \boldsymbol{x}^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \neq 0\right) > 0. \tag{20}$$

The following conditions (c3) – (c5) obviously hold for the distribution functions $F_p(z)$ under consideration and their densities

$$f_p(z) = (1-p)\,\boldsymbol{I}(0 < z \leq 1) + p\boldsymbol{I}(1 < z < 2), \qquad z \in \mathbb{R}. \tag{21}$$

(c3) Distributions functions $F_p(z)$ are continuous in both arguments $p \in (0,1)$ and $z \in (0,\infty)$. Moreover, for each $p \in (0,1)$

$$\int_{-\infty}^{+\infty} |z|\, f_p(z)\, \mathrm{d}z = \frac{1}{2} + p < \infty. \tag{22}$$

(c4) Distributions functions F_p, $p \in (0,1)$ are increasing on interval $[0,2] \subseteq \mathbb{R}$ in the strict sense

$$F_p(z_1) < F_p(z_2) \ \text{ for } z_1 < z_2 \text{ from } [0,2] \tag{23}$$

and constant on the complement $\mathbb{R} - [0,2]$.

(c5) Distributions functions F_p, $p \in (0,1)$ are stochastically ordered in the sense that for every $0 < p_1 < p_2 < 1$ and $z \in \mathbb{R}$ it holds $F_{p_1}(z) \geq F_{p_2}(z)$ where

$$F_{p_1}(z) > F_{p_2}(z) \ \text{ if } z \in [0,2]. \tag{24}$$

The present conditions (c1) – (c5) imply the assumptions (E1+), (E2), (EM1), (EM2) and (M1) – (M4) of Theorem 2 and Lemmas 8 and 9 in Liese and Vajda [12]. For a detailed proof of this assertion we refer to Section 3 of Hobza et al [7]. We shall check that in our model hold also the following less evident conditions of consistency and asymptotic normality.

(c6) For every $0 < p_1 < p_2 < 1$ there exists $a > 0$ such that the densities (21) and the median function $m(p)$ satisfy the condition

$$\Lambda(a) \equiv \inf_{|y| \leq a} \left(\inf_{p_1 \leq p \leq p_2} f_p(m(p) + y) \right) > 0. \tag{25}$$

(c7) The quantile function $m(p)$ is differentiable on $(0,1)$ and the derivative $m'(p)$ is locally Lipschitz in the sense that for every $p_0 \in (0,1)$

$$|m'(p) - m'(p_0)| \leq 2\,|p - p_0|\,.$$

(c8) The densities (21) satisfy for every $0 < p_1 < p_2 < 1$ the condition

$$\lim_{y \to 0} \sup_{p_1 \le p \le p_2} \left| f_p\left(m\left(p\right) + y\right) - f_p\left(m\left(p\right)\right) \right| = 0. \tag{26}$$

It is not difficult to establish, in a similar way to that in Lemma 2.1 in Hobza et al. [9], that in the present model the conditions (c6) – (c8) hold.

The median function $m\left(p\right)$ of (15) is bounded on $[0,1]$. By Lemma 8 in Liese and Vajda [12], this means that the sufficient condition of Lemma 9 ibid. reduces to (20) assumed in (c2). Hence, by Theorem 2 and Lemmas 8, 9 in Liese and Vajda [12], under (c1) – (c5) our *Med-estimator* $\widehat{\boldsymbol{\beta}}_n$ consistently estimates the true $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ provided the measure Q of (c1) defines the function

$$\boldsymbol{m}\left(\boldsymbol{\beta}\right) = \int_{\mathbb{R}} \int_{\mathcal{X}} \left| y - \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) \right| \mathrm{d}F_{\pi(\boldsymbol{x}^T\boldsymbol{\beta})}\left(y\right) \mathrm{d}Q\left(\boldsymbol{x}\right) \quad \text{for} \quad \varphi(t) = m(\pi(t)) \tag{27}$$

of variable $\boldsymbol{\beta} \in \mathbb{R}^d$ satisfying for every $\varepsilon > 0$ the condition

$$\inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \ge \varepsilon} \boldsymbol{m}\left(\boldsymbol{\beta}\right) > \boldsymbol{m}\left(\boldsymbol{\beta}_0\right) \tag{28}$$

of identifiability of true parameters $\boldsymbol{\beta}_0$. This important fact will be used in the proof of the following theorem.

Theorem 2.1. If the regressors of the model under consideration satisfy (c1), (c2) then the *Med-estimator* $\widehat{\boldsymbol{\beta}}_n$ consistently estimates the model parameters $\boldsymbol{\beta}_0$.

Proof. By what was said above, (c1) – (c8) hold. It suffices to prove that then (28) holds as well. Put for φ of (27)

$$\Delta = \Delta\left(\boldsymbol{x}, \boldsymbol{\beta}\right) = \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right) - \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) \tag{29}$$

and

$$\eta = \xi - \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right),$$

being ξ a random variable with density function $f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}(y)$ defined in (21). Then the density of η evaluated at point z is

$$g_{\boldsymbol{x}}\left(z\right) = f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}\left(z + \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)\right), \qquad z \in \mathbb{R},$$

and

$$\boldsymbol{m}\left(\boldsymbol{\beta}\right) - \boldsymbol{m}\left(\boldsymbol{\beta}_0\right) = \int_{\mathcal{X}} \left[w\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right) \right] \mathrm{d}Q\left(\boldsymbol{x}\right) \tag{30}$$

for

$$\begin{aligned} w\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) &= E\left| \xi - \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) \right| \\ &= E\left| \eta + \Delta\left(\boldsymbol{x}, \boldsymbol{\beta}\right) \right| \qquad \text{(cf. (29))}. \end{aligned}$$

The difference

$$w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) = E\left(|\eta + \Delta(\boldsymbol{x},\boldsymbol{\beta})| - |\eta|\right) \tag{31}$$

will be estimated by using the generalized Taylor formula

$$|\eta + \Delta| - |\eta| = D^+|\eta|\,\Delta + \mathcal{R}(\eta,\Delta) \tag{32}$$

valid for all real Δ where

$$D^+|z| = \boldsymbol{I}(0 \le z < \infty) - \boldsymbol{I}(-\infty < z < 0) \tag{33}$$

is the right-hand derivative of the function $|z|$ for $z \in \mathbb{R}$ and

$$\mathcal{R}(z,\Delta) = \begin{cases} 2(\Delta + z)\cdot \boldsymbol{I}(-\Delta < z < 0), & \text{for } \Delta > 0; \\ -2(\Delta + z)\cdot \boldsymbol{I}(0 < z < -\Delta), & \text{for } \Delta < 0, \end{cases}$$

is a remainder in the formula (32). This follows from the generalized Taylor expansion of arbitrary convex function established in (3.3) of Liese and Vajda [13] and the formula (3.6) ibid. for the remainder. Since $\mathrm{med}(\eta) = 0$, it holds $ED^+|\eta| = 0$. Therefore, from (31) and (32), we get for $\Delta > 0$

$$\begin{aligned} w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) &= E\,\mathcal{R}(\eta,\Delta) \\ &= \int 2(z+\Delta)\,\boldsymbol{I}(-\Delta < z < 0)\,g_{\boldsymbol{x}}(z)\,\mathrm{d}z \\ &= 2\int_0^\Delta (\Delta - z)\,g_{\boldsymbol{x}}(-z)\,\mathrm{d}z \\ &= 2\int_0^\Delta (\Delta - z)\,f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) - z\right)\mathrm{d}z. \end{aligned}$$

It is not difficult to see, using the corresponding form of the remainder $\mathcal{R}(z,\Delta)$, that the same formula is obtained also for $\Delta < 0$. Since $\mathcal{X} \subset \mathbb{R}^d$ is bounded, the values

$$p_1 = \inf_{\boldsymbol{x}\in\mathcal{X}} \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) \qquad \text{and} \qquad p_2 = \sup_{\boldsymbol{x}\in\mathcal{X}} \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)$$

are bounded away from 0 and 1. Thus, taking into account that $\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) = m\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)$, we see from (c6) that we can find $a > 0$ such that

$$\inf_{|z|\le a}\inf_{\boldsymbol{x}\in\mathcal{X}} f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) - z\right) \ge \Lambda(a) > 0.$$

This implies that if $0 < b < a$ then for every $|\Delta(\boldsymbol{x},\boldsymbol{\beta})| > b$ it holds

$$w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) \ge b^2\Lambda(a).$$

Hence, by (30), for every $0 < b < a$ we get

$$\boldsymbol{m}(\boldsymbol{\beta}) - \boldsymbol{m}(\boldsymbol{\beta}_0) \ge b^2\Lambda(a)\,Q\left(\mathcal{X}_{b,\boldsymbol{\beta}}\right) \tag{34}$$

for the subset of regressors

$$\mathcal{X}_{b,\boldsymbol{\beta}} = \{\, \boldsymbol{x} \in \mathcal{X} : |\Delta\,(\, \boldsymbol{x}, \boldsymbol{\beta})| \geq b\}\,.$$

By (c2), the smallest eigenvalue $\lambda\,(\boldsymbol{\Sigma})$ of the matrix (19) is positive. Further, for every $\tau > 0$

$$\begin{aligned}\lambda\,(\boldsymbol{\Sigma})\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2 &\leq (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\,\boldsymbol{\Sigma}\,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\ &= \int_{\mathcal{X}} \left(\boldsymbol{x}^T\,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right)^2\,\mathrm{d}Q\,(\boldsymbol{x})\\ &\leq \|\mathcal{X}\|\,.\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2\,Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right)+\tau^2\end{aligned}$$

where $\|\mathcal{X}\|$ stands for $\max\|\boldsymbol{x}\|$ on $\mathcal{X}$ and $\mathcal{X}^0_{\tau,\boldsymbol{\beta}} = \left\{\, \boldsymbol{x} \in \mathcal{X} : \left|\boldsymbol{x}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right| > \tau\right\}$. From here we see that for all $\varepsilon > 0$ and all sufficiently small $\tau > 0$

$$\psi\,(\tau,\varepsilon) \equiv \inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq\varepsilon} Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right) > 0. \tag{35}$$

It follows from the strict monotonicity and continuity of the functions $m(p)$ and $\pi(t)$ that $\varphi\,(t)$ of (27) is strictly monotone and continuous on $\mathbb{R}$. Therefore the function

$$\phi\,(\tau) \equiv \inf_{\substack{|t|\leq\|\mathcal{X}\|\cdot\|\boldsymbol{\beta}_0\|\\ |s-t|\geq\tau}} |\varphi\,(s)-\varphi\,(t)|$$

is positive in the domain $\tau > 0$ and, obviously,

$$\mathcal{X}_{\phi(\tau),\boldsymbol{\beta}} \supseteq \mathcal{X}^0_{\tau,\boldsymbol{\beta}}.$$

Since $\varphi\,(t)$ is continuous, it holds $\phi\,(\tau) < a$ for all sufficiently small $\tau > 0$. Consequently (34) implies for any $\varepsilon > 0$

$$\begin{aligned}\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq\varepsilon}\left[\boldsymbol{m}\,(\boldsymbol{\beta})-\boldsymbol{m}\,(\boldsymbol{\beta}_0)\right] &\geq \phi\,(\tau)^2\Lambda\,(a)\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq\varepsilon} Q\left(\mathcal{X}_{\phi(\tau),\boldsymbol{\beta}}\right)\\ &\geq \phi\,(\tau)^2\Lambda\,(a)\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq\varepsilon} Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right)\\ &= \phi\,(\tau)^2\Lambda\,(a)\,\psi\,(\tau,\varepsilon)\,.\end{aligned}$$

By (35), the last product is positive which proves the desired relation (28). □

Theorem 2.2. Let in the model under consideration the derivative $g'(\pi)$ of the link function be bounded away from zero on $(0,1)$ and the regressors satisfy (c1), (c2). If the limit matrix $\mathcal{Q}$ in (38) is positive definite then the *Med-estimator* $\widehat{\boldsymbol{\beta}}_n$ of the model parameters $\boldsymbol{\beta}_0$ is asymptotically normal in the sense that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) \xrightarrow[n\to\infty]{\mathcal{L}} N\left(\mathbf{0},\ \mathcal{Q}^{-1}\Sigma\mathcal{Q}^{-1}\right) \tag{36}$$

for Σ given by (37) and $\mathcal{Q}$ given by (38).

P r o o f. The function φ of (27) is continuously differentiable with the derivative $\varphi'(t) = m'(\pi(t))\pi'(t)$. From now on we shall assume that the derivative $g'(\pi)$ of the link function is bounded away from zero on $(0,1)$. This is equivalent to the assumption that the derivative $\pi'(t)$ is bounded on $\mathbb{R}$. Since $m'(p)$ is bounded on $(0,1)$, this implies that $\varphi'(t)$ is bounded on $\mathbb{R}$. Let us introduce the notation

$$\begin{aligned} \Delta_i(\boldsymbol{\beta}) &= \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) - \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right), \qquad \boldsymbol{\beta} \in \mathbb{R}^d, \\ \eta_i &= \xi_i - \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right), \quad i = 1, 2, \ldots, \end{aligned}$$

where ξ_i is a random variable with density function $f_{\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)}(y)$. Therefore

$$\widetilde{f}_i(z) =_{\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)} \left(z + \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right), \qquad z \in \mathbb{R},$$

is the probability density function of η_i. The functions $\Delta_i(\boldsymbol{\beta})$ are continuously differentiable on $\mathbb{R}^d$ with gradients

$$grad\,(\Delta_i(\boldsymbol{\beta})) = -\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\boldsymbol{x}_i.$$

Therefore the linear term $\mathcal{L}_n(h)$ considered in (2.3) of Liese and Vajda [14] is given here by

$$\mathcal{L}_n(\boldsymbol{h}) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D^+|\eta_i|\,\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\boldsymbol{x}_i^T\boldsymbol{h}, \qquad \boldsymbol{h} \in \mathbb{R}^d,$$

where $D^+|z|$ denotes the right-hand derivative (33). Since $E\,D^+|\eta_i| = 0$, the variance of $\mathcal{L}_n(\boldsymbol{h})$ is $\boldsymbol{h}^T\Sigma_n\boldsymbol{h}$ for the matrix given in accordance with (2.5) of Liese and Vajda [14] by

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} E\left(D^+|\eta_i|\right)^2\left(\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)^2\boldsymbol{x}_i\boldsymbol{x}_i^T.$$

But $E\left(D^+|\eta_i|\right)^2 = 1$ so that we can write the matrix Σ_n in the integral form

$$\Sigma_n = \int_{\mathcal{X}} \left(\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)^2\boldsymbol{x}^T\boldsymbol{x}\,\mathrm{d}Q_n(\boldsymbol{x})$$

where Q_n is the empirical measure from (c1). Since $\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)$ is continuous and bounded on $\mathcal{X}$, it holds

$$\lim_{n\to\infty}\Sigma_n = \Sigma \equiv \int_{\mathcal{X}} \left(\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)^2\boldsymbol{x}^T\boldsymbol{x}\,\mathrm{d}Q(\boldsymbol{x}) \tag{37}$$

where Q is the limit measure from (c1).

The next step is evaluation of the matrices

$$\mathcal{Q}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(0)\,\nabla\varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\left(\nabla\varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right)^T$$

where $g_i(t)$ denote derivatives of the functions $G_i(t) = E\mathcal{D}|\eta_i + t|$ of variable $t \in \mathbb{R}$ introduced on p. 467 in Liese and Vajda [13]. By the definition of $D^+|z|$ in (33), for $\pi_i = \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$ and $\varphi_i = \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$

$$\begin{aligned} G_i(t) &= EI(\eta_i + t > 0) - EI(\eta_i + t \le 0) \\ &= EI(\xi_i > \varphi_i - t) - EI(\xi_i \le \varphi_i - t) \\ &= 1 - 2F_{\pi_i}(\varphi_i - t). \end{aligned}$$

Thus $g_i(t) = 2f_{\pi_i}(\varphi_i - t)$ and

$$g_i(0) = 2f_{\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)}\left(\varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right).$$

Therefore the matrices $\mathcal{Q}_n$ may be represented as the integrals

$$\mathcal{Q}_n = 2\int_{\mathcal{X}} f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)\left(\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)^2 \boldsymbol{x}^T\boldsymbol{x}\,\mathrm{d}Q_n(\boldsymbol{x}).$$

Since $\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$ is continuous and bounded on $\mathcal{X}$ and, by (c8), the function

$$f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right) = f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}\left(m\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)\right)$$

is continuous and bounded on $\mathcal{X}$ too, it holds

$$\lim_{n\to\infty} \mathcal{Q}_n = \mathcal{Q} \equiv 2\int_{\mathcal{X}} f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)\left(\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)^2 \boldsymbol{x}^T\boldsymbol{x}\,\mathrm{d}Q(\boldsymbol{x}). \tag{38}$$

Finally, $D^+\rho(\eta_i) = D^+|\eta_i|$ is in the present situation bounded and $\nabla f_i(\boldsymbol{\beta}_0) = grad(\Delta_i(\boldsymbol{\beta}_0)) = -\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\boldsymbol{x}_i$ is bounded uniformly for all possible $\boldsymbol{x}_i \in \mathcal{X}$. Consequently the Liapunov condition (2.6) of Liese and Vajda [14] holds. Similarly, one can verify that the conditions (C3), (C4) of Liese and Vajda [13] as well as (2.39), (2.40) ibid. hold. Thus, by Lemma 3 in Liese and Vajda [13], (C5) and (C6) ibid. hold too. However, this allows to go a step further and conclude that if (c1), (c2) hold then all assumptions of Theorem 1 in Liese and Vajda [14] are satisfied. That theorem implies the enunciated result. □

3. ASYMPTOTIC DISTRIBUTION FOR SOME GENERALIZED LINEAR MODELS

In this section we shall obtain the expression of matrices Σ and $\mathcal{Q}$ for the generalized linear models considered in Section 1.

3.1. Logistic regression model

Let us consider the logistic regression model of (3). In this model we get from (27)

$$\varphi(t) = \begin{cases} \dfrac{3}{2} - \dfrac{e^{-t}}{2} & \text{if} \quad t \ge 0 \\ \dfrac{1}{2} + \dfrac{e^{t}}{2} & \text{if} \quad t < 0. \end{cases}$$

It is clear that

$$\varphi'(t) = \frac{e^{-|t|}}{2} \quad \text{if} \quad t \in \mathbb{R}$$

and

$$\begin{aligned} f_{\pi(t)}(\varphi(t)) &= \pi(t) \vee (1 - \pi(t)) \\ &= \frac{1}{1+e^{-t}} \vee \frac{1}{1+e^{t}} = \frac{1}{1+e^{-|t|}} = \frac{e^{|t|}}{1+e^{|t|}}. \end{aligned}$$

Therefore,

$$f_{\pi(t)}(\varphi(t))(\varphi'(t))^2 = \frac{e^{-|t|}}{4\left(1+e^{|t|}\right)}$$

and one obtains from (37) and (38)

$$\Sigma = \frac{1}{4}\int_{\mathcal{X}} e^{-2|\boldsymbol{x}^T\boldsymbol{\beta}_0|}\boldsymbol{x}\boldsymbol{x}^T \, \mathrm{d}Q(\boldsymbol{x}) \tag{39}$$

and

$$\mathcal{Q} = \frac{1}{2}\int_{\mathcal{X}} \frac{e^{-|\boldsymbol{x}^T\boldsymbol{\beta}_0|}}{1+e^{|\boldsymbol{x}^T\boldsymbol{\beta}_0|}} \; \boldsymbol{x}\boldsymbol{x}^T \, \mathrm{d}Q(\boldsymbol{x}). \tag{40}$$

Thus in the logistic regression model Theorem 2.2 holds with Σ and $\mathcal{Q}$ given by (39) and (40).

For more details about this model as well as the expressions of Σ and $\mathcal{Q}$ for the univariate logistic regression model see Hobza et al. [9].

3.2. Probit model

Let us now consider the probit model of (4). In this model (27) implies

$$\varphi(t) = \begin{cases} 2 - \dfrac{1}{2\Phi(t)} & \text{if} \quad t \geq 0 \\ \dfrac{1}{2}\dfrac{1}{\Phi(-t)} & \text{if} \quad t < 0\,. \end{cases}$$

Therefore

$$\varphi'(t) = \frac{1}{\sqrt{8\pi}} \cdot \frac{e^{-t^2}}{\Phi^2(|t|)} \quad \text{if} \quad t \in \mathbb{R}.$$

Further

$$f_{\pi(t)}(\varphi(t)) = \pi(t) \vee (1 - \pi(t)) = \Phi(t) \vee \Phi(-t) = \Phi(|t|)\,.$$

Consequently,

$$f_{\pi(t)}(\varphi(t))(\varphi'(t))^2 = \frac{1}{8\pi} \cdot \frac{e^{-t^2}}{\Phi^3(|t|)}$$

and one obtains from (37) and (38)

$$\Sigma = \frac{1}{8\pi}\int_{\mathcal{X}} \frac{e^{-(\boldsymbol{x}^T\boldsymbol{\beta}_0)^2}}{\Phi^4\left(|\boldsymbol{x}^T\boldsymbol{\beta}_0|\right)}\boldsymbol{x}\boldsymbol{x}^T \, \mathrm{d}Q(\boldsymbol{x}) \tag{41}$$

and

$$\mathcal{Q} = \frac{1}{4\pi}\int_{\mathcal{X}} \frac{e^{-(\boldsymbol{x}^T\boldsymbol{\beta}_0)^2}}{\Phi^3\left(|\boldsymbol{x}^T\boldsymbol{\beta}_0|\right)}\boldsymbol{x}\boldsymbol{x}^T \,\mathrm{d}Q(\boldsymbol{x})\,. \tag{42}$$

Thus in the probit model Theorem 2.2 holds with Σ and $\mathcal{Q}$ given by (41) and (42).

3.3. Log-Log model

Now we consider the log – log model of (5). In this model (27) implies

$$\varphi(t) = \begin{cases} \frac{1}{2}\cdot\frac{e^{e^t}}{e^{e^t}-1} & \text{if} \quad t \geq \ln(\ln 2) \\ 2 - \frac{1}{2}e^{e^t} & \text{if} \quad t < \ln(\ln 2)\,. \end{cases}$$

Therefore

$$\varphi'(t) = \begin{cases} -\frac{1}{2}\cdot\frac{e^{t+e^t}}{(e^{e^t}-1)^2} & \text{if} \quad t \geq \ln(\ln 2) \\ -\frac{1}{2}e^{t+e^t} & \text{if} \quad t < \ln(\ln 2)\,. \end{cases}$$

In the same way as in the previous models we get

$$f_{\pi(t)}\left(\varphi(t)\right) = \begin{cases} 1 - e^{-e^t} & \text{if} \quad t \geq \ln(\ln 2) \\ e^{-e^t} & \text{if} \quad t < \ln(\ln 2)\,. \end{cases}$$

Consequently,

$$f_{\pi(t)}\left(\varphi(t)\right)\left(\varphi'(t)\right)^2 = \begin{cases} \frac{1}{4}\,e^{2t+e^t}\left(e^{e^t}-1\right)^{-3} & \text{if} \quad t \geq \ln(\ln 2) \\ \frac{1}{4}\,e^{2t+e^t} & \text{if} \quad t < \ln(\ln 2) \end{cases}$$

and one obtains from (37) and (38)

$$\begin{aligned} \Sigma &= \frac{1}{4}\int_{\mathcal{X}} \exp\left(2\boldsymbol{x}^T\boldsymbol{\beta}_0 + 2\exp\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)\cdot\Big[\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0 < \ln\ln 2\right) \\ &+ \left(\exp(\exp\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)) - 1\right)^{-4}\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0 \geq \ln\ln 2\right)\Big]\boldsymbol{x}\boldsymbol{x}^T\,\mathrm{d}Q(\boldsymbol{x}) \end{aligned} \tag{43}$$

and

$$\begin{aligned} \mathcal{Q} &= \frac{1}{2}\int_{\mathcal{X}} \exp\left(2\boldsymbol{x}^T\boldsymbol{\beta}_0 + \exp\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)\cdot\Big[\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0 < \ln\ln 2\right) \\ &+ \left(\exp(\exp\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)) - 1\right)^{-3}\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0 \geq \ln\ln 2\right)\Big]\boldsymbol{x}\boldsymbol{x}^T\,\mathrm{d}Q(\boldsymbol{x})\,. \end{aligned} \tag{44}$$

Thus in the log-log model Theorem 2.2 holds with Σ and $\mathcal{Q}$ given by (43) and (44).

3.4. Complementary Log-Log model

We consider the complementary log – log model of (6). In this model (27) implies

$$\varphi(t)=\begin{cases} 2-\dfrac{1}{2}\cdot\dfrac{e^{e^t}}{e^{e^t}-1} & \text{if} \quad t\geq \ln(\ln 2) \\ \dfrac{1}{2}e^{e^t} & \text{if} \quad t<\ln(\ln 2)\,. \end{cases}$$

Therefore

$$\varphi'(t)=\begin{cases} \dfrac{1}{2}\cdot\dfrac{e^{t+e^t}}{(e^{e^t}-1)^2} & \text{if} \quad t\geq \ln(\ln 2) \\ \dfrac{1}{2}e^{t+e^t} & \text{if} \quad t<\ln(\ln 2)\,. \end{cases}$$

Further, we get

$$f_{\pi(t)}(\varphi(t))=\begin{cases} 1-e^{-e^t} & \text{if} \quad t\geq \ln(\ln 2) \\ e^{-e^t} & \text{if} \quad t<\ln(\ln 2)\,. \end{cases}$$

From the last formulas it follows that in the complementary log-log model Theorem 2.2 holds with Σ and $\mathcal{Q}$ given by the same formulas (43) and (44) as in the case of the log-log model.

3.5. Scobit model

Let us study the Scobit model of (7). In this model it holds

$$\varphi(t)=\begin{cases} \dfrac{3}{2}-\dfrac{1}{2}\cdot\dfrac{1}{(1+e^t)^\lambda-1} & \text{if} \quad t\geq \ln(2^{1/\lambda}-1) \\ \dfrac{1}{2}(1+e^t)^\lambda & \text{if} \quad t<\ln(2^{1/\lambda}-1)\,. \end{cases}$$

Therefore

$$\varphi'(t)=\begin{cases} \dfrac{\lambda}{2}(1+e^t)^{\lambda-1}e^t\left[(1+e^t)^\lambda-1\right]^{-2} & \text{if} \quad t\geq \ln(2^{1/\lambda}-1) \\ \dfrac{\lambda}{2}(1+e^t)^{\lambda-1}e^t & \text{if} \quad t<\ln(2^{1/\lambda}-1)\,. \end{cases}$$

In the same way as in the previous models we get

$$f_{\pi(t)}(\varphi(t))=\begin{cases} 1-(1+e^t)^{-\lambda} & \text{if} \quad t\geq \ln(2^{1/\lambda}-1) \\ (1+e^t)^{-\lambda} & \text{if} \quad t<\ln(2^{1/\lambda}-1)\,. \end{cases}$$

Consequently,

$$f_{\pi(t)}(\varphi(t))\left(\varphi'(t)\right)^2=\begin{cases} \dfrac{\lambda^2}{4}(1+e^t)^{\lambda-2}e^{2t}\left[(1+e^t)^\lambda-1\right]^{-3} & \text{if} \quad t\geq \ln(2^{1/\lambda}-1) \\ \dfrac{\lambda^2}{4}(1+e^t)^{\lambda-2}e^{2t} & \text{if} \quad t<\ln(2^{1/\lambda}-1)\,, \end{cases}$$

and one obtains from (37) and (38)

$$\begin{aligned}\Sigma &= \frac{\lambda^2}{4}\int_{\mathcal{X}} \left(1+e^{\boldsymbol{x}^T\boldsymbol{\beta}_0}\right)^{2\lambda-2} e^{2\boldsymbol{x}^T\boldsymbol{\beta}_0}\cdot\Big[\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0<\ln(2^{1/\lambda}-1)\right)\\ &+ \left(\left(1+e^{\boldsymbol{x}^T\boldsymbol{\beta}_0}\right)^{\lambda}-1\right)^{-4}\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\geq\ln(2^{1/\lambda}-1)\right)\Big]\boldsymbol{x}\boldsymbol{x}^T\,\mathrm{d}Q(\boldsymbol{x}) \qquad (45)\end{aligned}$$

and

$$\begin{aligned}\mathcal{Q} &= \frac{\lambda^2}{2}\int_{\mathcal{X}} \left(1+e^{\boldsymbol{x}^T\boldsymbol{\beta}_0}\right)^{\lambda-2} e^{2\boldsymbol{x}^T\boldsymbol{\beta}_0}\cdot\Big[\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0<\ln(2^{1/\lambda}-1)\right)\\ &+ \left(\left(1+e^{\boldsymbol{x}^T\boldsymbol{\beta}_0}\right)^{\lambda}-1\right)^{-3}\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\geq\ln(2^{1/\lambda}-1)\right)\Big]\boldsymbol{x}\boldsymbol{x}^T\,\mathrm{d}Q(\boldsymbol{x})\,. \qquad (46)\end{aligned}$$

Thus in the Scobit model Theorem 2.2 holds with Σ and $\mathcal{Q}$ given by (45) and (46).

3.6. Power logit model

Finally, let us suppose the Power logit model of (8). In this model it holds

$$\varphi(t)=\begin{cases}2-\frac{1}{2}\left(1+e^{-t}\right)^{\lambda} & \text{if} \quad t\geq-\ln(2^{1/\lambda}-1)\\ \frac{1}{2}\frac{1}{1-(1+e^{-t})^{-\lambda}} & \text{if} \quad t<-\ln(2^{1/\lambda}-1)\,.\end{cases}$$

Therefore

$$\varphi'(t)=\begin{cases}\frac{\lambda}{2}(1+e^{-t})^{\lambda-1}e^{-t} & \text{if} \quad t\geq-\ln(2^{1/\lambda}-1)\\ \frac{\lambda}{2}(1+e^{-t})^{\lambda-1}e^{-t}\left[(1+e^{-t})^{\lambda}-1\right]^{-2} & \text{if} \quad t<-\ln(2^{1/\lambda}-1)\,.\end{cases}$$

In the same way as in the previous models we get

$$f_{\pi(t)}\left(\varphi(t)\right)=\begin{cases}(1+e^{-t})^{-\lambda} & \text{if} \quad t\geq-\ln(2^{1/\lambda}-1)\\ 1-(1+e^{-t})^{-\lambda} & \text{if} \quad t<-\ln(2^{1/\lambda}-1)\,.\end{cases}$$

Consequently,

$$f_{\pi(t)}\left(\varphi(t)\right)\left(\varphi'(t)\right)^2=\begin{cases}\frac{\lambda^2}{4}(1+e^{-t})^{\lambda-2}e^{-2t} & \text{if } t\geq-\ln(2^{1/\lambda}-1)\\ \frac{\lambda^2}{4}(1+e^{-t})^{\lambda-2}e^{-2t}\left[(1+e^{-t})^{\lambda}-1\right]^{-3} & \text{if } t<-\ln(2^{1/\lambda}-1)\,,\end{cases}$$

and one obtains from (37) and (38)

$$\begin{aligned}\Sigma &= \frac{\lambda^2}{4}\int_{\mathcal{X}} \left(1+e^{-\boldsymbol{x}^T\boldsymbol{\beta}_0}\right)^{2\lambda-2} e^{-2\boldsymbol{x}^T\boldsymbol{\beta}_0}\cdot\Big[\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\geq-\ln(2^{1/\lambda}-1)\right)\\ &+ \left(\left(1+e^{-\boldsymbol{x}^T\boldsymbol{\beta}_0}\right)^{\lambda}-1\right)^{-4}\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0<-\ln(2^{1/\lambda}-1)\right)\Big]\boldsymbol{x}\boldsymbol{x}^T\,\mathrm{d}Q(\boldsymbol{x}) \qquad (47)\end{aligned}$$

and

$$\mathcal{Q} = \frac{\lambda^2}{2}\int_{\mathcal{X}} \left(1+e^{-\boldsymbol{x}^T\boldsymbol{\beta}_0}\right)^{\lambda-2} e^{-2\boldsymbol{x}^T\boldsymbol{\beta}_0} \cdot \Big[\boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0 \geq -\ln(2^{1/\lambda}-1)\right)$$
$$+ \left(\left(1+e^{-\boldsymbol{x}^T\boldsymbol{\beta}_0}\right)^{\lambda}-1\right)^{-3} \boldsymbol{I}\left(\boldsymbol{x}^T\boldsymbol{\beta}_0 < -\ln(2^{1/\lambda}-1)\right)\Big]\boldsymbol{x}\boldsymbol{x}^T\,\mathrm{d}Q(\boldsymbol{x}). \quad (48)$$

Thus in the Scobit model Theorem 2.2 holds with Σ and $\mathcal{Q}$ given by (47) and (48).

4. ROBUSTNESS

The application of median estimator to generalized linear models with binary responses was motivated by its familiar robustness in continuous regression models. So there was a hope it will be robust too. The *Med-estimator* is in this respect compared with several robust estimators known from the previous literature. Let us start with description of the compared estimators.

- **Morgenthaler estimator**

  The first robust estimator included in the simulation study was proposed by Morgenthaler [18] and is defined as the solution of the equations

$$\sum_{i=1}^{n} \boldsymbol{\psi}\left(Y_i, \boldsymbol{x}_i, \boldsymbol{\beta}\right) = \mathbf{0} \quad (49)$$

  for the function $\boldsymbol{\psi} : \mathbb{R}^{2d+1} \to \mathbb{R}^d$ given by the formula

$$\boldsymbol{\psi}\left(Y, \boldsymbol{x}, \boldsymbol{\beta}\right) = \sqrt{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\left(1-\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)}\left(Y - \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)\boldsymbol{x}. \quad (50)$$

  This estimator is called briefly *Morg-estimator* in the sequel.

- **Bianco–Yohai estimator**

  The second one is the M-estimator introduced by Bianco and Yohai [2] and defined as minimizer

$$\boldsymbol{\beta}_n = \arg\min \sum_{i=1}^{n} \phi\left(Y_i, \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right) \quad (51)$$

  for

$$\phi\left(Y_i, \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right) = \varrho\left(d_i\left(\boldsymbol{\beta}\right)\right) + \varrho_0\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right) \quad (52)$$

  where $d_i\left(\boldsymbol{\beta}\right)$ are the deviances (10) of individual observations Y_i, $\varrho\left(t\right)$ is bounded function specified by

$$\varrho\left(0\right) = 0 \quad \text{and} \quad \varrho'\left(t\right) = \left(1-t\right)\boldsymbol{I}\left(0 < t < 1\right) \quad (53)$$

  and the compensator function ϱ_0 is of the form

$$\varrho_0\left(p\right) = \varrho_1\left(p\right) + \varrho_1\left(1-p\right) \quad (54)$$

for ϱ_1 depending on ϱ by the formula

$$\varrho_1(p) = \int_0^p \varrho'(-\ln t)\,\mathrm{d}t, \qquad p \in (0,1). \tag{55}$$

This M-estimator is called briefly *BY-estimator* in the sequel.

- **Croux–Haesbroeck estimator**

  Croux and Haesbroeck [3] proposed an alternative estimator from the general Bianco–Yohai class defined by the formulas (51), (52) obtained for modified function $\varrho(t)$ satisfying

  $$\varrho(0) = 0 \quad \text{and} \quad \varrho'(t) = e^{-\sqrt{1/2}}\boldsymbol{I}(0 < t < 1/2) + e^{-\sqrt{t}}\boldsymbol{I}(t \geq 1/2). \tag{56}$$

  This particular M-estimator is called *CH-estimator* in the sequel

- **Enhanced median estimator**

  As illustrated in Example 2.1 of Hobza et al. [9], application of the L_1- estimators (13) in discrete statistical models with observations Y_i, $1 \leq i \leq n$, statistically smoothed into the continuous form (11) is usually accompanied by a loss of efficiency achievable in the original discrete models. This subefficiency can be suppressed to some extent by the method of enhancing introduced in Hobza et al. [8, 9]. It consists in replacing the set of statistically smoothed data $Z_i = Y_i + U_i$, $1 \leq i \leq n$ by the expanded set obtained by considering for $k > 1$ the matrix of data

  $$Z_{ij} = Y_i + U_{ij}, \qquad 1 \leq i \leq n, \qquad 1 \leq j \leq k, \tag{57}$$

  where U_{ij} are $U(0,1)$-distributed and mutually as well as on $Y_1, \ldots, Y_n$ independent random variables, and applying the L_1-estimator to this expanded set. If a method of processing the data $Z_1, \ldots, Z_n$ is statistically optimal in an appropriate sense, then its performance cannot be improved by expanding the sufficient statistic $(Z_1, \ldots, Z_n)$. On the other hand, if the method is suboptimal, like for example the median estimator of the above mentioned Example 2.1, then its performance can be improved by using the expanded data set (57). In some cases the efficiency of the median estimator can even reach asymptotically the efficiency of the MLE estimates. For more details concerning enhancement we refer to Hobza et al. [8, 9].

  In the following we shall call the estimator obtained by applying the Med-estimator (13) to the expanded data set (57) by *k-Med-estimator.*

An extensive simulation study comparing properties of *Med-estimator, Morg-estimator, BY-estimator* and *CH-estimator* under the logistic regression model was done in Hobza et al. [8] and reported also in Hobza et al. [9]. The results show the robustness of the median estimators by demonstrating their low sensitivity to high leverage points and also by demonstrating that they outperform the above mentioned classical robust estimators in certain special situations (e. g. heavy contaminations and large sample sizes). The conclusions are based partly on simulations in the models used in the previous literature for mutual comparison of various estimators in logistic regression.

In the present paper we present a new simulation study in order to see if the conclusions concerning robustness obtained in the above cited work for logistic regression models remains valid also in other GLM's. We are going to present the results for probit models but they continue being valid in other models.

4.1. Simulation experiment

The robustness is compared by means of simulated performances of all selected estimators in the probit models $Be\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)$ ε-contaminated at the levels $0 \le \varepsilon \le 0.3$ by the alternative data source $Be\left(1-\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)$, or contaminated at the same levels ε by the leverage points from probit models $Be(\pi(\widetilde{\boldsymbol{x}}^T\boldsymbol{\beta}_0))$ with strongly distorted regressors $\widetilde{\boldsymbol{x}}$ at the place of $\boldsymbol{x}$.

The first data source is very similar to the one used previously by Bianco and Yohai [2] to demonstrate experimentally the robustness of their *BY-estimator* in logistic regression which was modified slightly to the probit regression. The simulated data $Y_1, \ldots, Y_n$ are generated by the contaminated probit source

$$Y_i \sim (1-\varepsilon)\,Be\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right) + \varepsilon Be\left(1-\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right), \tag{58}$$

where $\pi(t)$ is the standard normal distribution function (cf. (4)), $\boldsymbol{x}_i$ are the concrete regressors

$$\boldsymbol{x}_i = (x_{i0} \equiv 1,\ x_{i1} \sim N(0,1))^T \tag{59}$$

and $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (-1.87,\ 2.0)^T$ are the true parameters leading to the probability

$$\Pr(Y_i = 1) \equiv E\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) = 0.2\,. \tag{60}$$

In the second case of the leverage points the simulated data $Y_1, \ldots, Y_n$ are generated by the source

$$Y_i \sim (1-\varepsilon)\,Be\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right) + \varepsilon Be\left(\pi\left(\widetilde{\boldsymbol{x}}_i^T\boldsymbol{\beta}_0\right)\right) \tag{61}$$

with the same regressors $\boldsymbol{x}_i$ and true parameters $\boldsymbol{\beta}_0$ as used in (58), but with the regressors $\widetilde{\boldsymbol{x}}_i$ different, given by the formula

$$\widetilde{\boldsymbol{x}}_i = \left(1, \widetilde{x}_{i1} = \beta_{00} + 3sign\left[-\frac{\beta_{00}}{\beta_{01}} - x_{i1}\right]\beta_{01}\right), \quad x_{i1} \sim N(0,1), \tag{62}$$

and characterized by the property

$$\pi\left(\ \boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) > 1/2 \quad \text{implies} \quad \pi\left(\widetilde{\boldsymbol{x}}_i^T\boldsymbol{\beta}_0\right) \approx 0$$

and

$$\pi\left(\ \boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) \le 1/2 \quad \text{implies} \quad \pi\left(\widetilde{\boldsymbol{x}}_i^T\boldsymbol{\beta}_0\right) \approx 1.$$

In all cases the results presented in Tables 1−6 are based on 1000 simulated realizations $\widetilde{\boldsymbol{\beta}}_n^{(l)} = (\widetilde{\beta}_{n0}^{(l)}, \widetilde{\beta}_{n1}^{(l)})^T$ of an estimator $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of true parameters $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T$. These realizations have been used to evaluate the *mean absolute errors*

$$\text{MAE} = \frac{1}{2000}\sum_{l=1}^{1000}\left(\left|\widetilde{\beta}_{n0}^{(l)} - \beta_{00}\right| + \left|\widetilde{\beta}_{n1}^{(l)} - \beta_{01}\right|\right) \tag{63}$$

and also the *rejection rates* RR specifying for each selected estimator the percentage of the data vectors $Y_1, \ldots, Y_n$ rejected during evaluation of the desired 1000 realizations. The data vector is rejected and replaced by a new independent realization if numerical evaluation of one of the assumed estimators fails. For the sake of completeness, we have included in the tables also results for the MLE.

The estimates $\widetilde{\boldsymbol{\beta}}_n^{(l)} = (\widetilde{\beta}_{n0}^{(l)}, \widetilde{\beta}_{n1}^{(l)})^T$ were evaluated numerically in accordance with the definition of each corresponding estimator given above, using the iteration procedures presented in the *IMSL C Numerical Libraries*, version 3.0. The minimization of a function of two variables uses there a quasi-Newton method (for details see Appendix A of Dennis and Schnabel [4]), and systems of equations are solved using a modified Powell hybrid algorithm (for further description see Moré et al. [17]). The initial iteration seeds for the MLE $\beta_n = (\beta_{n0}, \beta_{n1})^T$ were the true parameters $\beta_0 = (\beta_{00}, \beta_{01})^T$ and the initial iteration seeds for all the remaining estimates $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ were the MLE's $\boldsymbol{\beta}_n = (\beta_{n0}, \beta_{n1})^T$.

We observe that the results presented in the Tables 1 and 2 follow basically the same pattern. From the first uncontaminated ($\varepsilon = 0$) sector of the tables one can see that the best behavior has the MLE estimator except for the case of $n = 50$ observations where the *BY-estimator* is the preferred one .

From the ε-contaminated sectors of the Tables 1 and 2 it follows that for small sample sizes ($n = 50, 100$) the best results are obtained alternately for *Morg-* and *BY-estimators.* However the main message of the tables is that for larger sample sizes ($n = 500, 1000$) our *Med-estimator* or its k-enhanced version better resists to higher levels of contamination or distortion by leverage points than the remaining four estimators known from the previous literature.

From the sectors corresponding to $n = 500, 1000$ and $\varepsilon = 0, 0.05$ one can also observe how the method of k-enhancing can improve the behavior of the *Med-estimator*. For higher levels of contamination this is not true but on the other hand the enhancing does not make the behavior of the *Med-estimator* much worse so that for these sample sizes (around $n \approx 500$) the use of the k-enhancing for $k \approx 10$ can be recommended.

The above introduced MAE represents in some sense an overall measure of estimates precision. In order to distinguish a bias and a variability of an estimator and to single out the performance of an estimator of slope and of an estimator of intercept we present in Tables 3 − 6 also the following measures based on the above described simulation experiment. Namely, the estimated bias (BIAS)

$$\mathrm{BIAS}(i) = \frac{1}{1000} \sum_{l=1}^{1000} \left(\widetilde{\beta}_{ni}^{(l)} - \beta_{0i} \right), \qquad i = 0, 1,$$

and the mean absolute variability (MAV)

$$\mathrm{MAV}(i) = \frac{1}{1000} \sum_{l=1}^{1000} \left| \widetilde{\beta}_{ni}^{(l)} - \mathrm{mean}(i) \right|, \qquad i = 0, 1,$$

where *mean*(i) was obtained for each of the assumed estimators by the formula

$$\text{mean}(i) = \frac{1}{10000} \sum_{l=1}^{10000} \widetilde{\beta}_{ni}^{(l)}, \qquad i = 0, 1,$$

using 10 000 independent realizations of the estimates $\widetilde{\beta}_{ni}^{(l)}$, $i = 0, 1$, different from that ones used for the calculation of MAV.

From Tables 3 and 5 one can see that as soon as there is some contamination the *Med-estimator* have the smallest absolute values of BIAS for larger sample sizes $n = 500, 1000$. For smaller sample sizes $= 50$ and $n = 100$ the same happens for levels of contamination $\varepsilon \geq 0.2$ or $\varepsilon \geq 0.1$, respectively. Concerning the measure of variability MAV presented in Tables 4 and 6 the best in this respect is in almost all cases the MLE estimator as expected. The variability of the *Med-estimator* and its k-enhanced versions is significantly bigger particularly for small sample sizes $n = 50, 100$ but for larger sample sizes ($n = 500, 1000$) and higher level of contamination the variability of the *10-Med-estimator* is comparable with variability of the other robust estimators as can be seen e.g. from Tables 4 and 6 and sectors corresponding to $n = 500, 1000$ and $\varepsilon \geq 0.1$. Let us remind that at the same time it shows smaller BIAS that the rest of the compared robust estimators.

From the tables can be seen also that there is no much difference between the behavior of the estimator of intercept ($\widetilde{\beta}_{n0}$) and the behavior of the estimator of slope ($\widetilde{\beta}_{n1}$).

As a conclusion from all the presented simulation results we can say that for larger sample sizes ($n \approx 500$) and higher level of contamination ($\varepsilon \geq 0.1$) the *Med-estimator* and its k-enhanced versions have shown to be good competitors to the other compared robust estimators.

		$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
	MLE	0.818	2	**0.406**	0	**0.158**	0	**0.107**	0
	Morg	0.947	4	0.537	0	0.196	0	0.128	0
	BY	**0.739**	2	0.414	0	0.173	0	0.116	0
	CH	1.552	6	0.711	1	0.237	0	0.150	0
0	Med	3.480	14	2.188	6	0.771	0	0.346	0
	5-Med	2.688	19	1.494	7	0.393	0	0.232	0
	10-Med	2.773	19	1.577	8	0.365	0	0.213	0
	MLE	0.915	0	0.830	0	0.876	0	0.903	0
	Morg	0.998	1	**0.623**	0	0.478	0	0.475	0
	BY	**0.860**	0	0.684	0	0.670	0	0.681	0
	CH	1.448	4	0.716	0	**0.349**	0	0.329	0
0.05	Med	3.037	9	1.963	5	0.631	0	0.395	0
	5-Med	2.906	14	1.621	7	0.423	0	0.321	0
	10-Med	2.855	16	1.757	10	0.380	0	**0.316**	0
	MLE	1.119	0	1.123	0	1.149	0	1.169	0
	Morg	1.076	1	0.912	0	0.909	0	0.919	0
	BY	**1.049**	0	1.017	0	1.027	0	1.044	0
	CH	1.320	2	**0.869**	0	0.710	0	0.709	0
0.1	Med	2.592	7	1.763	4	0.770	0	0.612	0
	5-Med	2.423	13	1.536	5	0.717	0	**0.609**	0
	10-Med	2.748	14	1.604	10	**0.688**	0	0.613	0
	MLE	1.442	0	1.431	0	1.439	0	1.449	0
	Morg	**1.395**	0	1.377	0	1.380	0	1.392	0
	BY	1.419	0	1.398	0	1.399	0	1.410	0
	CH	1.431	0	**1.334**	0	1.365	0	1.379	0
0.2	Med	2.032	5	1.503	2	1.220	0	**1.198**	0
	5-Med	1.918	6	1.616	3	1.207	0	1.209	0
	10-Med	2.159	7	1.778	4	**1.205**	0	1.211	0
	MLE	1.633	0	1.631	0	1.629	0	1.638	0
	Morg	1.622	0	1.620	0	1.616	0	1.626	0
	BY	1.626	0	1.622	0	1.617	0	1.627	0
	CH	**1.621**	0	1.619	0	1.618	0	1.628	0
0.3	Med	1.898	2	**1.592**	1	**1.538**	0	**1.563**	0
	5-Med	1.789	2	1.643	1	1.550	0	1.568	0
	10-Med	1.791	2	1.563	0	1.553	0	1.568	0

Tab. 1. MAE and RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n$ of the true parameter $\boldsymbol{\beta}_0$ in the ε-contaminated probit regression model (58). (The achieved minima are printed bold.)

		$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
ε	(β_{n1}, β_{n2})	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.818	2	**0.406**	0	**0.158**	0	**0.107**	0
	Morg	0.947	4	0.537	0	0.196	0	0.128	0
	BY	**0.739**	2	0.414	0	0.173	0	0.116	0
	CH	1.552	6	0.711	1	0.237	0	0.150	0
	Med	3.480	14	2.188	6	0.771	0	0.346	0
	5-Med	2.688	19	1.494	7	0.393	0	0.232	0
	10-Med	2.773	19	1.577	8	0.365	0	0.213	0
0.05	MLE	0.917	0	0.845	0	0.884	0	0.919	0
	Morg	0.957	1	**0.608**	0	0.516	0	0.533	0
	BY	**0.856**	0	0.699	0	0.689	0	0.716	0
	CH	1.327	3	0.696	0	**0.394**	0	**0.394**	0
	Med	3.155	9	1.934	6	0.598	0	0.435	0
	5-Med	2.932	13	1.581	5	0.426	0	0.403	0
	10-Med	2.884	15	1.816	8	0.418	0	0.396	0
0.1	MLE	1.125	0	1.133	0	1.159	0	1.180	0
	Morg	**1.061**	0	0.932	0	0.947	0	0.967	0
	BY	1.067	0	1.035	0	1.047	0	1.069	0
	CH	1.221	2	**0.832**	0	0.783	0	0.799	0
	Med	2.371	6	1.630	3	0.759	0	**0.756**	0
	5-Med	2.224	12	1.347	4	**0.726**	0	0.758	0
	10-Med	2.509	13	1.420	7	0.730	0	0.761	0
0.2	MLE	1.448	0	1.439	0	1.451	0	1.462	0
	Morg	**1.417**	0	1.390	0	1.402	0	1.414	0
	BY	1.428	0	1.409	0	1.417	0	1.429	0
	CH	1.392	0	**1.358**	0	1.391	0	1.406	0
	Med	2.123	3	1.527	2	**1.273**	0	**1.294**	0
	5-Med	1.768	4	1.359	1	1.285	0	1.302	0
	10-Med	1.734	5	1.360	1	1.288	0	1.303	0
0.3	MLE	1.643	0	1.641	0	1.644	0	1.651	0
	Morg	**1.636**	0	**1.633**	0	1.634	0	1.642	0
	BY	1.639	0	1.635	0	1.635	0	1.643	0
	CH	1.655	0	**1.633**	0	1.635	0	1.644	0
	Med	1.904	1	1.642	1	**1.586**	0	**1.604**	0
	5-Med	1.694	1	1.576	0	1.595	0	1.608	0
	10-Med	1.676	2	1.585	0	1.597	0	1.609	0

Tab. 2. MAE and RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n$ of the true parameter $\boldsymbol{\beta}_0$ in the probit regression model (61) ε-contaminated by leverage points. (The achieved minima are printed bold.)

		$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
ε	$\widetilde{\beta}$	BIAS		BIAS		BIAS		BIAS	
		$\widetilde{\beta}_{n0}$	$\widetilde{\beta}_{n1}$	$\widetilde{\beta}_{n0}$	$\widetilde{\beta}_{n1}$	$\widetilde{\beta}_{n0}$	$\widetilde{\beta}_{n1}$	$\widetilde{\beta}_{n0}$	$\widetilde{\beta}_{n1}$
	MLE	-0.408	0.531	-0.184	0.215	**-0.031**	0.039	**-0.018**	0.020
	Morg	-0.528	0.590	-0.235	0.268	-0.040	0.048	-0.022	0.024
	BY	**-0.293**	1.171	**-0.153**	0.452	**-0.031**	0.077	**-0.018**	0.034
	CH	-1.091	**0.361**	-0.401	**0.178**	-0.066	**0.038**	-0.031	**0.019**
0	Med	-2.986	2.862	-1.638	1.788	-0.376	0.424	-0.136	0.152
	5-Med	-2.457	2.411	-1.244	1.385	-0.177	0.204	-0.068	0.075
	10-Med	-2.583	2.606	-1.247	1.375	-0.164	0.189	-0.061	0.068
	MLE	0.593	-0.815	0.759	-0.885	0.809	-0.943	0.827	-0.986
	Morg	**0.159**	**-0.200**	0.345	-0.372	0.456	-0.497	0.460	-0.506
	BY	0.512	0.535	0.609	**-0.016**	0.633	-0.317	0.645	-0.338
	CH	-0.499	-0.664	**0.021**	-0.678	0.295	-0.706	0.314	-0.730
0.05	Med	-1.825	1.527	-1.019	1.122	**-0.015**	**0.022**	**0.202**	**-0.213**
	5-Med	-1.806	1.726	-0.778	0.856	0.174	-0.182	0.256	-0.273
	10-Med	-2.024	2.020	-0.985	1.073	0.181	-0.190	0.268	-0.286
	MLE	0.951	-1.268	1.049	-1.216	1.066	-1.232	1.074	-1.264
	Morg	0.705	-0.878	0.823	-0.907	0.862	-0.948	0.869	-0.968
	BY	0.900	**-0.254**	0.968	-0.537	0.968	-0.726	0.978	-0.739
	CH	**0.183**	-1.155	0.502	-1.088	0.673	-1.083	0.680	-1.110
0.1	Med	-1.029	0.745	**-0.406**	**0.446**	**0.374**	**-0.393**	**0.539**	**-0.580**
	5-Med	-1.056	0.965	**-0.406**	0.465	0.515	-0.552	0.583	-0.628
	10-Med	-1.266	1.228	-0.487	0.547	0.537	-0.577	0.589	-0.634
	MLE	1.282	-1.601	1.346	-1.528	1.350	-1.526	1.353	-1.548
	Morg	1.248	-1.528	1.306	-1.459	1.306	-1.450	1.313	-1.474
	BY	1.271	-1.397	1.323	-1.380	1.322	-1.431	1.326	-1.460
	CH	1.156	-1.568	1.246	-1.485	1.293	-1.473	1.303	-1.495
0.2	Med	**0.187**	-0.418	**0.486**	**-0.511**	**1.083**	**-1.167**	**1.141**	**-1.238**
	5-Med	0.341	-0.456	0.704	-0.746	1.149	-1.240	1.160	-1.260
	10-Med	0.190	**-0.319**	0.681	-0.728	1.154	-1.246	1.163	-1.263
	MLE	1.504	-1.761	1.546	-1.714	1.547	-1.713	1.548	-1.728
	Morg	1.498	-1.747	1.538	-1.699	1.539	-1.696	1.541	-1.712
	BY	1.501	-1.724	1.540	-1.698	1.540	-1.698	1.542	-1.715
	CH	1.484	-1.753	1.537	-1.701	1.540	-1.698	1.542	-1.713
0.3	Med	**0.888**	**-1.119**	**1.228**	**-1.320**	**1.484**	**-1.606**	**1.495**	**-1.628**
	5-Med	1.151	-1.307	1.394	-1.507	1.492	-1.616	1.499	-1.633
	10-Med	1.177	-1.339	1.413	-1.530	1.493	-1.618	1.499	-1.634

Tab. 3. BIAS for selected estimators $\widetilde{\boldsymbol{\beta}}_n$ of the true parameter $\boldsymbol{\beta}_0$ in the ε-contaminated probit regression model (58). (The achieved minima of absolute values are printed bold.) .

ε	$\widetilde{\beta}$	$n=50$ MAV $\widetilde{\beta}_{n0}$	$n=50$ MAV $\widetilde{\beta}_{n1}$	$n=100$ MAV $\widetilde{\beta}_{n0}$	$n=100$ MAV $\widetilde{\beta}_{n1}$	$n=500$ MAV $\widetilde{\beta}_{n0}$	$n=500$ MAV $\widetilde{\beta}_{n1}$	$n=1000$ MAV $\widetilde{\beta}_{n0}$	$n=1000$ MAV $\widetilde{\beta}_{n1}$
0	MLE	**0.730**	0.999	**0.392**	**0.460**	**0.147**	**0.170**	**0.098**	**0.117**
	Morg	0.947	1.216	0.547	0.612	0.181	0.210	0.118	0.140
	BY	0.654	2.074	0.402	0.836	0.160	0.253	0.106	0.163
	CH	1.856	**0.920**	0.763	0.465	0.221	0.186	0.140	0.127
	Med	4.551	4.477	2.584	2.859	0.783	0.860	0.343	0.382
	5-Med	3.539	3.598	1.847	2.079	0.371	0.420	0.218	0.250
	10-Med	3.655	3.812	1.890	2.114	0.341	0.391	0.200	0.232
0.05	MLE	**0.376**	**0.579**	**0.223**	**0.292**	**0.082**	**0.106**	**0.055**	**0.068**
	Morg	0.783	1.041	0.416	0.497	0.137	0.159	0.099	0.117
	BY	0.418	1.806	0.260	0.758	0.100	0.201	0.069	0.144
	CH	1.576	0.614	0.660	0.328	0.181	0.119	0.123	0.084
	Med	3.674	3.788	2.280	2.567	0.603	0.673	0.302	0.338
	5-Med	3.669	3.703	1.873	2.045	0.331	0.374	0.188	0.214
	10-Med	3.620	3.801	2.133	2.320	0.286	0.324	0.172	0.195
0.1	MLE	**0.226**	**0.340**	**0.149**	**0.190**	**0.061**	**0.074**	**0.042**	**0.053**
	Morg	0.499	0.697	0.276	0.342	0.099	0.119	0.069	0.086
	BY	0.253	1.329	0.173	0.639	0.072	0.177	0.051	0.126
	CH	1.083	0.379	0.563	0.220	0.156	0.087	0.104	0.062
	Med	2.934	3.141	1.874	2.062	0.530	0.584	0.256	0.284
	5-Med	2.791	3.013	1.674	1.872	0.351	0.387	0.153	0.178
	10-Med	3.237	3.514	1.806	1.976	0.293	0.325	0.140	0.163
0.2	MLE	**0.158**	**0.220**	**0.113**	**0.132**	**0.049**	**0.059**	**0.034**	**0.039**
	Morg	0.181	0.279	0.125	0.163	0.056	0.070	0.039	0.050
	BY	0.162	0.534	0.117	0.253	0.052	0.085	0.036	0.061
	CH	0.333	0.238	0.185	0.144	0.065	0.064	0.045	0.044
	Med	1.676	1.994	1.087	1.256	0.258	0.299	0.120	0.139
	5-Med	1.462	1.827	1.012	1.172	0.119	0.140	0.078	0.095
	10-Med	1.804	2.152	1.181	1.322	0.103	0.123	0.073	0.089
0.3	MLE	**0.150**	**0.189**	**0.105**	**0.120**	**0.045**	**0.051**	**0.033**	**0.036**
	Morg	0.152	0.201	0.107	0.128	0.046	0.054	**0.033**	0.038
	BY	**0.150**	0.229	**0.105**	0.130	0.046	0.054	**0.033**	0.038
	CH	0.167	0.195	0.108	0.125	0.046	0.053	**0.033**	0.038
	Med	1.005	1.202	0.443	0.544	0.079	0.093	0.052	0.066
	5-Med	0.625	0.845	0.302	0.374	0.061	0.072	0.041	0.050
	10-Med	0.576	0.831	0.196	0.250	0.057	0.067	0.040	0.048

Tab. 4. MAV for selected estimators $\widetilde{\boldsymbol{\beta}}_n$ of the true parameter $\boldsymbol{\beta}_0$ in the ε-contaminated probit regression model (58). (The achieved minima are printed bold.)

ε	$\widetilde{\beta}$	$n=50$ BIAS $\widetilde{\beta}_{n0}$	$n=50$ BIAS $\widetilde{\beta}_{n1}$	$n=100$ BIAS $\widetilde{\beta}_{n0}$	$n=100$ BIAS $\widetilde{\beta}_{n1}$	$n=500$ BIAS $\widetilde{\beta}_{n0}$	$n=500$ BIAS $\widetilde{\beta}_{n1}$	$n=1000$ BIAS $\widetilde{\beta}_{n0}$	$n=1000$ BIAS $\widetilde{\beta}_{n1}$
0	MLE	-0.408	0.531	-0.184	0.215	**-0.031**	0.039	**-0.018**	0.020
	Morg	-0.528	0.590	-0.235	0.268	-0.040	0.048	-0.022	0.024
	BY	**-0.293**	1.171	**-0.153**	0.452	**-0.031**	0.077	**-0.018**	0.034
	CH	-1.091	**0.361**	-0.401	**0.178**	-0.066	**0.038**	-0.031	**0.019**
	Med	-2.986	2.862	-1.638	1.788	-0.376	0.424	-0.136	0.152
	5-Med	-2.457	2.411	-1.244	1.385	-0.177	0.204	-0.068	0.075
	10-Med	-2.583	2.606	-1.247	1.375	-0.164	0.189	-0.061	0.068
0.05	MLE	0.613	-0.823	0.767	-0.891	0.818	-0.945	0.836	-0.991
	Morg	**0.226**	**-0.267**	0.394	-0.423	0.495	-0.532	0.504	-0.551
	BY	0.541	0.384	0.629	**-0.123**	0.654	-0.372	0.668	-0.402
	CH	-0.353	-0.686	**0.128**	-0.698	0.354	-0.721	0.374	-0.752
	Med	-1.762	1.479	-0.793	0.876	**0.133**	**-0.124**	**0.314**	**-0.333**
	5-Med	-1.648	1.577	-0.626	0.714	0.277	-0.283	0.358	-0.381
	10-Med	-1.842	1.849	-0.784	0.877	0.291	-0.298	0.361	-0.385
0.1	MLE	0.971	-1.263	1.058	-1.222	1.079	-1.234	1.087	-1.268
	Morg	0.771	-0.931	0.874	-0.962	0.904	-0.984	0.912	-1.009
	BY	0.927	**-0.445**	0.988	-0.680	0.992	-0.806	1.001	-0.822
	CH	**0.377**	-1.161	0.630	-1.108	0.756	-1.098	0.760	-1.129
	Med	-0.787	0.552	-0.193	0.263	**0.590**	**-0.615**	**0.696**	**-0.744**
	5-Med	-0.832	0.769	**-0.074**	**0.107**	0.692	-0.726	0.721	-0.773
	10-Med	-0.996	0.990	-0.134	0.176	0.702	-0.738	0.722	-0.775
0.2	MLE	1.307	-1.584	1.358	-1.537	1.372	-1.529	1.373	-1.551
	Morg	1.282	-1.524	1.327	-1.480	1.337	-1.464	1.341	-1.489
	BY	1.299	-1.416	1.340	-1.437	1.349	-1.451	1.352	-1.480
	CH	1.208	-1.555	1.296	-1.500	1.328	-1.483	1.335	-1.507
	Med	**0.353**	**-0.505**	**0.798**	**-0.854**	**1.224**	**-1.297**	**1.245**	**-1.341**
	5-Med	0.553	-0.630	0.995	-1.059	1.247	-1.324	1.255	-1.353
	10-Med	0.461	-0.523	0.987	-1.037	1.249	-1.326	1.256	-1.354
0.3	MLE	1.542	-1.740	1.560	-1.724	1.577	-1.713	1.574	-1.729
	Morg	1.538	-1.729	1.555	-1.713	1.571	-1.699	1.569	-1.717
	BY	1.540	-1.713	1.556	-1.713	1.571	-1.701	1.570	-1.719
	CH	1.530	-1.733	1.555	-1.715	1.572	-1.701	1.570	-1.717
	Med	**1.067**	**-1.196**	**1.365**	**-1.467**	**1.544**	**-1.637**	**1.548**	**-1.666**
	5-Med	1.305	-1.375	1.485	-1.609	1.550	-1.646	1.550	-1.669
	10-Med	1.350	-1.409	1.510	-1.632	1.550	-1.647	1.550	-1.669

Tab. 5. BIAS for selected estimators $\widetilde{\boldsymbol{\beta}}_n$ of the true parameter $\boldsymbol{\beta}_0$ in the probit regression model (61) ε-contaminated by leverage points. (The achieved minima of absolute values are printed bold.)

		$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
ε	$\widetilde{\beta}$	MAV		MAV		MAV		MAV	
		$\widetilde{\beta}_{n0}$	$\widetilde{\beta}_{n1}$	$\widetilde{\beta}_{n0}$	$\widetilde{\beta}_{n1}$	$\widetilde{\beta}_{n0}$	$\widetilde{\beta}_{n1}$	$\widetilde{\beta}_{n0}$	$\widetilde{\beta}_{n1}$
	MLE	**0.730**	0.999	**0.392**	**0.460**	**0.147**	**0.170**	**0.098**	**0.117**
	Morg	0.947	1.216	0.547	0.612	0.181	0.210	0.118	0.140
	BY	0.654	2.074	0.402	0.836	0.160	0.253	0.106	0.163
	CH	1.856	**0.920**	0.763	0.465	0.221	0.186	0.140	0.127
0	Med	4.551	4.477	2.584	2.859	0.783	0.860	0.343	0.382
	5-Med	3.539	3.598	1.847	2.079	0.371	0.420	0.218	0.250
	10-Med	3.655	3.812	1.890	2.114	0.341	0.391	0.200	0.232
	MLE	**0.363**	**0.563**	**0.217**	**0.280**	**0.081**	**0.105**	**0.059**	**0.075**
	Morg	0.701	0.970	0.374	0.443	0.133	0.157	0.093	0.112
	BY	0.392	1.626	0.241	0.692	0.097	0.200	0.070	0.133
	CH	1.380	0.582	0.610	0.299	0.175	0.118	0.112	0.086
0.05	Med	3.775	3.892	2.180	2.383	0.513	0.578	0.253	0.289
	5-Med	3.603	3.688	1.784	1.949	0.276	0.316	0.164	0.192
	10-Med	3.580	3.807	2.111	2.277	0.265	0.297	0.152	0.178
	MLE	**0.220**	**0.334**	**0.147**	**0.183**	**0.060**	**0.072**	**0.041**	**0.051**
	Morg	0.416	0.610	0.244	0.303	0.091	0.111	0.065	0.080
	BY	0.246	1.102	0.169	0.497	0.070	0.156	0.048	0.114
	CH	0.873	0.372	0.428	0.210	0.132	0.084	0.094	0.060
0.1	Med	2.639	2.818	1.664	1.832	0.334	0.385	0.196	0.224
	5-Med	2.536	2.767	1.307	1.495	0.185	0.222	0.116	0.141
	10-Med	2.932	3.160	1.449	1.582	0.166	0.198	0.106	0.130
	MLE	**0.162**	**0.217**	**0.113**	**0.130**	**0.048**	**0.056**	**0.035**	**0.040**
	Morg	0.186	0.270	0.126	0.158	0.053	0.067	0.038	0.048
	BY	0.164	0.371	0.117	0.211	0.049	0.078	0.036	0.055
	CH	0.243	0.229	0.159	0.142	0.058	0.061	0.042	0.043
0.2	Med	1.668	1.983	0.873	1.000	0.137	0.170	0.084	0.108
	5-Med	1.168	1.505	0.481	0.577	0.081	0.106	0.059	0.078
	10-Med	1.190	1.512	0.460	0.556	0.077	0.100	0.055	0.073
	MLE	**0.150**	**0.196**	**0.106**	**0.120**	**0.044**	**0.050**	**0.033**	**0.035**
	Morg	**0.150**	0.206	0.107	0.126	0.045	0.053	0.034	0.037
	BY	**0.150**	0.251	0.107	0.128	**0.044**	0.053	0.034	0.037
	CH	0.168	0.201	0.109	0.124	0.045	0.052	0.034	0.037
0.3	Med	0.849	1.141	0.331	0.444	0.069	0.090	0.050	0.062
	5-Med	0.399	0.634	0.145	0.199	0.053	0.071	0.040	0.049
	10-Med	0.327	0.595	0.129	0.179	0.051	0.067	0.038	0.048

Tab. 6. MAV for selected estimators $\widetilde{\boldsymbol{\beta}}_n$ of the true parameter $\boldsymbol{\beta}_0$ in the probit regression model (61) ε-contaminated by leverage points. (The achieved minima are printed bold.)

ACKNOWLEDGEMENT

This work was partially supported by the grants MTM 2009-10072, MTM 2009-06997, GAČR P202/10/0618 and MŠMT 6840770039.

(Received January 19, 2011)

REFERENCES

[1] G. Adimari and L. Ventura: Robust inference for generalized linear models with application to logistic regression. Statist. Probab. Lett. *55* (2001), 4, 413–419.

[2] A. M. Bianco and V. J. Yohai: Robust estimation in the logistic regression model. In: Robust Statistics, Data Analysis, and Computer Intensive Methods (Schloss Thurnau, 1994), pp. 17–34. Lecture Notes in Statist. *109* Springer, New York 1996.

[3] C. Croux and G. Haesbroeck: Implementing the Bianco and Yohai estimator for logistic regression. Computat. Statist. Data Anal. *44* (2003), 273–295.

[4] J. E. Dennis, Jr. and R. B. Schnabel: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs, New Jersey 1983.

[5] D. Gervini: Robust adaptive estimators for binary regression models. J. Statist. Plann. Inference *131* (2005), 297–311.

[6] F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti, and W. A. Stahel: Robust Statistics: The Approach Based on Influence Functions. Wiley, New York 1986.

[7] T. Hobza, L. Pardo, and I. Vajda: Median Estimators in Generalized Logistic Regression. Research Report DAR-UTIA 2005/40. Institute of Information Theory, Prague 2005 (available at http://dar.site.cas.cz/?publication=1007).

[8] T. Hobza, L. Pardo, and I. Vajda: Robust Median Estimators in Logistic Regression. Research Report DAR-UTIA 2006/31. Institute of Information Theory, Prague 2006 (available at http://dar.site.cas.cz/?publication=1089).

[9] T. Hobza, L. Pardo, and I. Vajda: Robust median estimator in logistic regression. J. Statist. Plann. Inference *138* (2008), 3822–3840.

[10] J. Jurečková and P. K. Sen: Robust Statistical Procedures. Wiley, New York 1996.

[11] N. Kordzakhia, G. D. Mishra, and L. Reiersølmoen: Robust estimation in the logistic regression model. J. Statist. Plann. Inference *98* (2001), 211–223.

[12] F. Liese and I. Vajda: M-estimators of structural parameters in pseudolinear models. Appl. Math. *44* (1999), 245–270.

[13] F. Liese and I. Vajda: A general asymptotic theory of M-estimators I. Math. Methods Statist. *12* (2003) 454–477.

[14] F. Liese and I. Vajda: A general asymptotic theory of M-estimators II. Math. Methods Statist. *13* (2004) 82–95.

[15] P. McCullagh and J. A. Nelder: Generalized Linear Models. Chapman and Hall, London 1989.

[16] M. L. Menéndez, L. Pardo, and M. C. Pardo: Preliminary test estimators and phi-divergence measures in generalized linear models with binary data. J. Multivariate Anal. *99* (2009), 10, 2265–2284.

[17] J. Moré, G. Burton, and H. Kenneth: User Guide for MINPACK-1. Argonne National Laboratory Report ANL-80-74, Argonne 1980.

[18] S. Morgenthaler: Least-absolute-deviations fits for generalized linear models. Biometrika *79* (1992), 747–754.

[19] J. Nagler: Scobit: An alternative estimator to logit and Probit. Amer. J. Political Sci. *38* (1994), 1, 230–255.

[20] D. Pregibon: Resistant lits for some commonly used logistic models with medical applications. Biometrics *38* (1982), 485–498.

[21] R. L. Prentice: A generalization of the probit and logit methods for dose-response curves. Biometrika *32* (1976), 761–768.

[22] P. J. Rousseeuw and A. Christmann: Robustness against separation and outliers in logistic regression. Comput. Statist. Data Anal. *43* (2003), 315–332.

[23] S. Zwanzig: On L_1-norm estimators in nonlinear regression and in nonlinear error-in-variables models. IMS Lect. Notes *31* (1997), 101–118.

Tomáš Hobza, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2. Czech Republic.
e-mail: hobza@fjfi.cvut.cz

Leandro Pardo, Universidad Complutense de Madrid, Departamento de Estadística e I.O. Plaza de Ciencias 3, 28040 Madrid. Spain.
e-mail: lpardo@mat.ucm.es

Igor Vajda, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: –